

CONSISTENCY OF DIRECT INTEGRAL ESTIMATOR FOR PARTIALLY OBSERVED SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS LINEAR IN THE PARAMETERS *

IVAN VUJAČIĆ¹ AND ITAI DATTNER²

Abstract. Dynamic systems are ubiquitous in nature and are used to model many processes in biology, chemistry, physics, medicine, and engineering. In particular, systems of ordinary differential equations are commonly used for the mathematical modelling of the rate of change of dynamic processes. In many practical applications, the process can only be partially measured, a fact that renders estimation of parameters of the system extremely challenging. Recently, a 'direct integral estimator' for partially observed systems of ordinary differential equations was introduced. The practical performance of the integral estimator was demonstrated, but its theoretical properties were not derived. In this paper we use the sieve framework to prove that the estimator is consistent.

1991 Mathematics Subject Classification. 62F12, 34A55.

1. INTRODUCTION

Mathematical models defined by a system of ordinary differential equations (ODEs) are commonly used for modelling dynamic processes. The process of interest is usually modelled by the system

$$\begin{cases} \mathbf{x}'(t) = \mathbf{F}(\mathbf{x}(t); \boldsymbol{\theta}), & t \in [0, T], \\ \mathbf{x}(0) = \boldsymbol{\xi}, \end{cases} \quad (1)$$

where $\mathbf{x}(t) \in \mathbb{R}^d$, $\boldsymbol{\xi} \in \Xi \subset \mathbb{R}^d$, and $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$.

Given the values of $\boldsymbol{\xi}$ and $\boldsymbol{\theta}$, we denote the solution of (1) by

$$\mathbf{x}(t) = \mathbf{x}(t; \boldsymbol{\theta}, \boldsymbol{\xi}), \quad t \in [0, T].$$

The aim is to estimate the unknown parameter $\boldsymbol{\theta}$ (and if necessary $\boldsymbol{\xi}$) from noisy observations

$$Y_j(t_i) = x_j(t_i; \boldsymbol{\theta}, \boldsymbol{\xi}) + \varepsilon_{j,i}, \quad i = 1, \dots, n, j = 1, \dots, r, \quad (2)$$

Keywords and phrases: Consistency, ordinary differential equation, nonparametric regression, sieve extremum estimators.

* This research is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO), and which is partly funded by Ministry of Economic Affairs. Part of this research was done during a visit of the second author to The Netherlands, supported by STAR Visitor Grant.

¹ Department of Mathematics, VU University Amsterdam, Room S3.30, De Boelelaan 1081a, 1081HV Amsterdam, The Netherlands

² Department of Statistics, University of Haifa, 199 Aba Khoushy Ave. Mount Carmel, Haifa 3498838, Israel

where $0 \leq t_1 < \dots < t_n = T < \infty$ and $\varepsilon_{j,i}$ is the unobserved measurement error for x_j at time t_i . When the number of measured states r is equal to d the system is fully observed, the case well studied in the literature. Partially observed systems, i.e. when $r < d$, are common in practice but much less studied. For an approach to estimation in this setting see [9], and [8] for its asymptotic analysis. Recently, [3] extended the direct integral estimator, defined for fully observed systems linear in functions of the parameters [4, 10], to partially observed ones. In the case of systems linear in the parameters, the ODE system studied in [4] has the form

$$\mathbf{F}(\mathbf{x}(t); \boldsymbol{\theta}) = \mathbf{g}(\mathbf{x}(t))\boldsymbol{\theta}, \quad (3)$$

where $\mathbf{g} : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times p}$ maps the d -dimensional column vector \mathbf{x} into a $d \times p$ matrix. The introduced estimator for fully observed systems, i.e. when $r = d$, is motivated by the system of integral equations

$$\mathbf{x}(t) = \boldsymbol{\xi} + \int_0^t \mathbf{g}(\mathbf{x}(s)) \, ds \boldsymbol{\theta}, \quad t \in [0, T], \quad (4)$$

which follows from (1) and (3) by integration. In view of (4), the estimators of the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ are obtained by minimizing

$$\int_0^T \|\widehat{\mathbf{x}}(t) - \boldsymbol{\xi} - \int_0^t \mathbf{g}(\widehat{\mathbf{x}}(s)) \, ds \boldsymbol{\theta}\|^2 \, dt$$

with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$, where $\widehat{\mathbf{x}}(t)$, $t \in [0, T]$, is a specific estimator of $\mathbf{x}(t; \boldsymbol{\theta}, \boldsymbol{\xi})$. Here and subsequently, $\|\cdot\|$ denotes the Euclidean norm. Since the objective function in the display above is quadratic in $(\boldsymbol{\theta}, \boldsymbol{\xi})$, it has a unique point of minimum $(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\xi}}_n)$ given by

$$\begin{aligned} \widehat{\boldsymbol{\xi}}_n &= (T\mathbf{I}_d - \widehat{\mathbf{A}}\widehat{\mathbf{B}}^{-1}\widehat{\mathbf{A}}^\top)^{-1} \int_0^T \{\mathbf{I}_d - \widehat{\mathbf{A}}\widehat{\mathbf{B}}^{-1}\widehat{\mathbf{G}}^\top(t)\} \widehat{\mathbf{x}}(t) \, dt, \\ \widehat{\boldsymbol{\theta}}_n &= \widehat{\mathbf{B}}^{-1} \int_0^T \widehat{\mathbf{G}}^\top(t) \{\widehat{\mathbf{x}}(t) - \widehat{\boldsymbol{\xi}}_n\} \, dt, \end{aligned} \quad (5)$$

where \mathbf{I}_d denotes the $d \times d$ identity matrix and

$$\begin{aligned} \widehat{\mathbf{G}}(t) &= \int_0^t \mathbf{g}(\widehat{\mathbf{x}}(s)) \, ds, \\ \widehat{\mathbf{A}} &= \int_0^T \widehat{\mathbf{G}}(t) \, dt, \\ \widehat{\mathbf{B}} &= \int_0^T \widehat{\mathbf{G}}^\top(t) \widehat{\mathbf{G}}(t) \, dt. \end{aligned} \quad (6)$$

If $\widehat{\mathbf{x}}(\cdot)$ is a consistent estimator of $\mathbf{x}(\cdot)$ in the sup norm then, under certain conditions, $(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\xi}}_n)$ is a consistent estimator of $(\boldsymbol{\theta}, \boldsymbol{\xi})$ [4].

We now describe the construction of the estimator for partially observed systems, i.e. when $r < d$, developed in [3]. Let \mathcal{M} and \mathcal{U} denote the sets of r -dimensional and $(d-r)$ -dimensional vector functions on $[0, T]$ that correspond to $\mathbf{m}(\cdot) = \mathbf{m}(\cdot, \boldsymbol{\theta}, \boldsymbol{\xi})$ and $\mathbf{u}^*(\cdot) = \mathbf{u}^*(\cdot, \boldsymbol{\theta}, \boldsymbol{\xi})$, the measured and unmeasured components, respectively. We first construct an estimator $\widehat{\mathbf{m}}_n(\cdot)$ of $\mathbf{m}(\cdot)$ using the observations (2) and for a given $\mathbf{u} \in \mathcal{U}$, in view of (6),

we define

$$\begin{aligned}\widehat{\mathbf{x}}_{\mathbf{u}}(t) &= (\widehat{\mathbf{m}}_n(t), \mathbf{u}(t)), \\ \widehat{\mathbf{G}}_{\mathbf{u}}(t) &= \int_0^t \mathbf{g}(\widehat{\mathbf{x}}_{\mathbf{u}}(s)) \, ds, \\ \widehat{\mathbf{A}}_{\mathbf{u}} &= \int_0^T \widehat{\mathbf{G}}_{\mathbf{u}}(t) \, dt, \\ \widehat{\mathbf{B}}_{\mathbf{u}} &= \int_0^T \widehat{\mathbf{G}}_{\mathbf{u}}^\top(t) \widehat{\mathbf{G}}_{\mathbf{u}}(t) \, dt.\end{aligned}\tag{7}$$

According to (5), the direct integral estimator based on $\widehat{\mathbf{x}}_{\mathbf{u}}(\cdot)$ is

$$\begin{aligned}\widehat{\boldsymbol{\xi}}_{\mathbf{u}} &= (T\mathbf{I}_d - \widehat{\mathbf{A}}_{\mathbf{u}}\widehat{\mathbf{B}}_{\mathbf{u}}^{-1}\widehat{\mathbf{A}}_{\mathbf{u}}^\top)^{-1} \int_0^T \{\mathbf{I}_d - \widehat{\mathbf{A}}_{\mathbf{u}}\widehat{\mathbf{B}}_{\mathbf{u}}^{-1}\widehat{\mathbf{G}}_{\mathbf{u}}^\top(t)\} \widehat{\mathbf{x}}_{\mathbf{u}}(t) \, dt, \\ \widehat{\boldsymbol{\theta}}_{\mathbf{u}} &= \widehat{\mathbf{B}}_{\mathbf{u}}^{-1} \int_0^T \widehat{\mathbf{G}}_{\mathbf{u}}^\top(t) \{\widehat{\mathbf{x}}_{\mathbf{u}}(t) - \widehat{\boldsymbol{\xi}}_{\mathbf{u}}\} \, dt.\end{aligned}\tag{8}$$

In case \mathbf{u}^* is measured, we can construct $\widehat{\mathbf{u}}_n(\cdot)$ and the above estimator reduces to the one from (5). Otherwise, for a subset $\mathcal{U}_n \subset \mathcal{U}$ let

$$\widehat{\mathbf{u}}_n := \operatorname{argmin}_{\mathbf{u} \in \mathcal{U}_n} M_n(\mathbf{u}),\tag{9}$$

where

$$M_n(\mathbf{u}) = \int_0^T \|\widehat{\mathbf{x}}_{\mathbf{u}}(t) - \widehat{\boldsymbol{\xi}}_{\mathbf{u}} - \widehat{\mathbf{G}}_{\mathbf{u}}(t)\widehat{\boldsymbol{\theta}}_{\mathbf{u}}\|^2 \, dt.\tag{10}$$

The estimators of initial value $\boldsymbol{\xi}$ and the parameter $\boldsymbol{\theta}$ are

$$\begin{aligned}\widehat{\boldsymbol{\xi}}_n &:= \widehat{\boldsymbol{\xi}}_{\widehat{\mathbf{u}}_n}, \\ \widehat{\boldsymbol{\theta}}_n &:= \widehat{\boldsymbol{\theta}}_{\widehat{\mathbf{u}}_n}.\end{aligned}\tag{11}$$

The aim of this paper is to establish consistency of $(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\xi}}_n)$ given by (11) under suitable assumptions.

As pointed out above, [4] studied fully observed systems and showed that if the estimator $(\widehat{\mathbf{m}}_n(\cdot), \widehat{\mathbf{u}}_n(\cdot))$ of $\mathbf{x}(\cdot)$ is consistent in the sup norm, then $(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\xi}}_n)$ is a consistent estimator of $(\boldsymbol{\theta}, \boldsymbol{\xi})$. However, in partially observed systems only \mathbf{m} is observed. Therefore, based on the data we are only able to construct consistent estimator $\widehat{\mathbf{m}}_n(\cdot)$ of $\mathbf{m}(\cdot)$. Consequently, our aim is to prove that consistent $\widehat{\mathbf{m}}_n(\cdot)$ gives rise to consistent estimator $\widehat{\mathbf{u}}_n(\cdot)$ of $\mathbf{u}^*(\cdot)$, defined in (9). To this end, define the deterministic counterparts of (7)-(8),

$$\begin{aligned}\mathbf{x}_{\mathbf{u}}(t) &= (\mathbf{m}(t), \mathbf{u}(t)), \\ \mathbf{G}_{\mathbf{u}}(t) &= \int_0^t \mathbf{g}(\mathbf{x}_{\mathbf{u}}(s)) \, ds, \\ \mathbf{A}_{\mathbf{u}} &= \int_0^T \mathbf{G}_{\mathbf{u}}(t) \, dt, \\ \mathbf{B}_{\mathbf{u}} &= \int_0^T \mathbf{G}_{\mathbf{u}}^\top(t) \mathbf{G}_{\mathbf{u}}(t) \, dt, \\ \boldsymbol{\xi}_{\mathbf{u}} &= (T\mathbf{I}_d - \mathbf{A}_{\mathbf{u}}\mathbf{B}_{\mathbf{u}}^{-1}\mathbf{A}_{\mathbf{u}}^\top)^{-1} \int_0^T \{\mathbf{I}_d - \mathbf{A}_{\mathbf{u}}\mathbf{B}_{\mathbf{u}}^{-1}\mathbf{G}_{\mathbf{u}}^\top(t)\} \mathbf{x}_{\mathbf{u}}(t) \, dt, \\ \boldsymbol{\theta}_{\mathbf{u}} &= \mathbf{B}_{\mathbf{u}}^{-1} \int_0^T \mathbf{G}_{\mathbf{u}}^\top(s) \{\mathbf{x}_{\mathbf{u}}(s) - \boldsymbol{\xi}_{\mathbf{u}}\} \, ds,\end{aligned}$$

and the asymptotic criterion

$$M(\mathbf{u}) = \int_0^T \|\mathbf{x}_{\mathbf{u}}(t) - \boldsymbol{\xi}_{\mathbf{u}} - \mathbf{G}_{\mathbf{u}}(t)\boldsymbol{\theta}_{\mathbf{u}}\|^2 \, dt,$$

corresponding to (10). Since $\widehat{\mathbf{u}}_n(\cdot)$ results from a minimization over sieves \mathcal{U}_n , it is a sieve extremum estimator [2]. Corollary 2.6 from [11, p. 467] provides the following conditions (c.f. Section 3.1, [2, p. 5589, 5590]) which are sufficient for consistency of $\widehat{\mathbf{u}}_n(\cdot)$.

- C1** $\mathcal{U}_n \subset \mathcal{U}_{n+1} \subset \mathcal{U}$ and for any $\mathbf{u} \in \mathcal{U}$ there exists $\pi_n \mathbf{u} \in \mathcal{U}_n$ such that $\|\mathbf{u} - \pi_n \mathbf{u}\|_\infty \rightarrow 0$ as $n \rightarrow \infty$.
- C2** \mathcal{U}_n is compact under $\|\cdot\|_\infty$.
- C3** Functional $M(\mathbf{u})$ is continuous at \mathbf{u}^* in \mathcal{U} under $\|\cdot\|_\infty$ and $M(\mathbf{u}^*) < +\infty$.
- C4** $M_n(\mathbf{u})$ is a measurable function of the data $\{Y_j(t_i)\}_{j,i}$ for all $\mathbf{u} \in \mathcal{U}_n$.
- C5** For any data $\{Y_j(t_i)\}_{j,i}$, $M_n(\mathbf{u})$ is lower semi-continuous on $\mathbf{u} \in \mathcal{U}_n$ under $\|\cdot\|_\infty$.
- C6** For all $\epsilon > 0$, $M(\mathbf{u}^*) < \inf_{\{\mathbf{u} \in \mathcal{U} : \|\mathbf{u} - \mathbf{u}^*\|_\infty \geq \epsilon\}} M(\mathbf{u})$.
- C7** $\sup_{\mathbf{u} \in \mathcal{U}_n} |M_n(\mathbf{u}) - M(\mathbf{u})| \xrightarrow{P} 0$ as $n \rightarrow \infty$.

Some additional notation: as in [5], for simplicity we omit the superscript $*$ in the outer probability P^* whenever an outer probability applies. The metric on \mathcal{U} and \mathcal{U}_n is the one induced by $\|\cdot\|_\infty$. For a matrix function $\mathbf{M} : [0, T] \rightarrow \mathbb{R}^{m \times p}$ we use the norm $\|\mathbf{M}\|_\infty = \sup_{t \in [0, T]} \|\mathbf{M}(t)\|$, where $\|\cdot\|$ is the Frobenius norm on $\mathbb{R}^{m \times p}$. The metric on the space of matrix functions on $[0, T]$ is the one induced by $\|\cdot\|_\infty$. $\widehat{\mathbf{M}}$ denotes the estimator of \mathbf{M} based on the data $\{Y_j(t_i)\}_{j,i}$.

In the next section we present an example of sieves \mathcal{U}_n that satisfy conditions C1 and C2. Consistency of $\widehat{\mathbf{u}}_n(\cdot)$ will be proven by verifying the conditions C3-C7. This in turn will imply consistency of $(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\xi}}_n)$.

The rest of the paper is organised as follows. In the next section we formulate the theorem dealing with consistency. The proofs are given in Section 3. The Appendix contains technical lemmas used in Section 3.

2. RESULTS

Consistency is established under the following assumptions:

Assumption 2.1.

- a) (*Existence and uniqueness*) For any $(\boldsymbol{\theta}, \boldsymbol{\xi}) \in \Theta \times \Xi$ there exists a unique solution $\mathbf{x}(\cdot; \boldsymbol{\theta}, \boldsymbol{\xi})$ of (1) on $[0, T]$.
- b) (*Identifiability*) For any $(\boldsymbol{\theta}, \boldsymbol{\xi}) \neq (\boldsymbol{\theta}', \boldsymbol{\xi}')$ it holds that $\mathbf{m}(\cdot; \boldsymbol{\theta}, \boldsymbol{\xi}) \neq \mathbf{m}(\cdot; \boldsymbol{\theta}', \boldsymbol{\xi}')$.
- c) (*Densness and compactness of sieves*) Sieves \mathcal{U}_n satisfy conditions C1 and C2.
- d) (*Compactness of function spaces*) $\mathcal{U} \subset \underbrace{C^1([0, T]) \times \cdots \times C^1([0, T])}_{d-r}$ and $\mathcal{M} \subset \underbrace{C^1([0, T]) \times \cdots \times C^1([0, T])}_r$

are compact under $\|\cdot\|_\infty$ and $\mathbf{m} \in \mathcal{M}$, $\mathbf{u}^* \in \mathcal{U}$.

Assumptions 2.1 a) and b) are usual in the estimation of parameters in ODE systems, see, for example, [8]. Assumption 2.1 b) is equivalent to: $\mathbf{m}(\cdot; \boldsymbol{\theta}, \boldsymbol{\xi}) = \mathbf{m}(\cdot; \boldsymbol{\theta}', \boldsymbol{\xi}') \Rightarrow (\boldsymbol{\theta}, \boldsymbol{\xi}) = (\boldsymbol{\theta}', \boldsymbol{\xi}')$. We do not require the converse implication because it is contained in Assumption 2.1 a). In other words,

Remark 2.2. Assumptions 2.1 a) and b) imply $\mathbf{m}(\cdot; \boldsymbol{\theta}, \boldsymbol{\xi}) = \mathbf{m}(\cdot; \boldsymbol{\theta}', \boldsymbol{\xi}') \iff (\boldsymbol{\theta}, \boldsymbol{\xi}) = (\boldsymbol{\theta}', \boldsymbol{\xi}')$.

Assumption 2.1 c) is necessary for consistency of the sieve extremum estimator $\widehat{\mathbf{u}}_n$; see previous section. We now give an example of sieves that satisfy this assumption. Define

$$\mathcal{U}_n = \{\mathbf{u} \in \mathcal{U} : u_j(t) = \sum_{k=1}^{K_{j,n}} \beta_{j,k} \phi_{j,k}(t), t \in [0, T], \beta_{j,k} \in \mathbb{R}; \sum_{j=1}^{d-r} \sum_{k=1}^{K_{j,n}} |\beta_{j,k}| \leq \Delta_n\},$$

where $\{\phi_{j,k}\}$ is a given sequence of basis functions such that $\|\phi_{j,k}\|_\infty < +\infty$, $j = 1, \dots, d-r$ and $k = 1, \dots, K_{j,n}$. For C1 and C2 to hold we can take, for example, $\phi_{j,k}$ to be cubic splines and $\min_{1 \leq j \leq d-r} (K_{j,n}, \Delta_n) \rightarrow \infty$ as $n \rightarrow \infty$. Indeed, Lemma 1 from [8] implies that the condition C1 is satisfied. The proof that C2 holds is the same like in [11, p. 471]. Here we assume that $(\widehat{\boldsymbol{\theta}}_{\mathbf{u}}, \widehat{\boldsymbol{\xi}}_{\mathbf{u}})$ and $(\boldsymbol{\theta}_{\mathbf{u}}, \boldsymbol{\xi}_{\mathbf{u}})$ are well-defined on \mathcal{U}_n , i.e., the matrix inverses that appear in their definition exist. Assumption 2.1 d) is a technical one; it is essential for our proof. We now formulate our main result.

Theorem 2.3. *Let the model be defined by (1),(3),(4) with the map $\mathbf{g} : \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^p$ continuous. Fix $\boldsymbol{\xi} \in \Xi$ and $\boldsymbol{\theta} \in \Theta$ and assume that $\mathbf{x}(\cdot) = \mathbf{x}(\cdot; \boldsymbol{\theta}, \boldsymbol{\xi})$ exists and is bounded on $[0, T]$, such that*

$$\|\mathbf{x}\|_\infty < \infty.$$

Assume that Assumption 2.1 holds. Let $\widehat{\mathbf{m}}_n(\cdot)$ be a consistent estimator of $\mathbf{m}(\cdot) = \mathbf{m}(\cdot; \boldsymbol{\theta}, \boldsymbol{\xi})$ in the supnorm, i.e.,

$$\|\widehat{\mathbf{m}}_n - \mathbf{m}\|_\infty \xrightarrow{P} 0.$$

Then the estimators $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\boldsymbol{\xi}}_n$ defined in (11) are consistent, i.e.,

$$(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\xi}}_n) \xrightarrow{P} (\boldsymbol{\theta}, \boldsymbol{\xi})$$

holds as $n \rightarrow \infty$.

3. PROOFS

Before proving the main result, we state a lemma that gives important asymptotic relationships which are used in the proof. Some of the results below implicitly use Lemma 4.3 of Appendix A.

Lemma 3.1. *Let \mathcal{U} and \mathcal{U}_n satisfy Assumption 2.1 c), d). Then as $n \rightarrow \infty$*

- (i) $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{x}}_{\mathbf{u}} - \mathbf{x}_{\mathbf{u}}\|_\infty = o_P(1)$, $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{x}_{\mathbf{u}}\|_\infty = O(1)$, $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{x}}_{\mathbf{u}}\|_\infty = O_P(1)$.
- (ii) $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{G}}_{\mathbf{u}} - \mathbf{G}_{\mathbf{u}}\|_\infty = o_P(1)$, $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{G}_{\mathbf{u}}\|_\infty = O(1)$, $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{G}}_{\mathbf{u}}\|_\infty = O_P(1)$.
- (iii) $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\boldsymbol{\xi}}_{\mathbf{u}} - \boldsymbol{\xi}_{\mathbf{u}}\| = o_P(1)$, $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\boldsymbol{\xi}_{\mathbf{u}}\| = O(1)$, $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\boldsymbol{\xi}}_{\mathbf{u}}\| = O_P(1)$.
- (iv) $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\boldsymbol{\theta}}_{\mathbf{u}} - \boldsymbol{\theta}_{\mathbf{u}}\| = o_P(1)$, $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\boldsymbol{\theta}_{\mathbf{u}}\| = O(1)$, $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\boldsymbol{\theta}}_{\mathbf{u}}\| = O_P(1)$.

Proof. In (i)-(iv) it suffices to prove only the first two statements of each, since they imply the third by using triangle inequality.

(i) For any $\mathbf{u} \in \mathcal{U}_n$, $\|\widehat{\mathbf{x}}_{\mathbf{u}} - \mathbf{x}_{\mathbf{u}}\|_\infty = \|\widehat{\mathbf{m}}_n - \mathbf{m}\|_\infty$ and hence the first assertion follows from consistency of $\widehat{\mathbf{m}}_n(\cdot)$. The second statement follows from the compactness of \mathcal{U}_n and boundedness of $\mathbf{m}(\cdot)$.

(ii) Introduce $K = \sup_{\mathbf{u} \in \mathcal{U}} \|\mathbf{x}_{\mathbf{u}}\|_\infty$, which by Assumption 2.1 d) is finite. Continuity of $\mathbf{g}(\cdot)$ on the closed compact ball $B_{K+1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \leq K+1\}$ implies its uniform continuity on B_{K+1} . Thus for $\epsilon > 0$ fixed there exists $\delta > 0$ such that for any $\mathbf{x}_1, \mathbf{x}_2 \in B_{K+1}$ if $\|\mathbf{x}_1 - \mathbf{x}_2\| < \delta$ then $\|\mathbf{g}(\mathbf{x}_1) - \mathbf{g}(\mathbf{x}_2)\| < \epsilon/T$. Now we show that

$$\|\widehat{\mathbf{m}}_n - \mathbf{m}\|_\infty < \min(1, \delta) \Rightarrow \sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{G}}_{\mathbf{u}} - \mathbf{G}_{\mathbf{u}}\|_\infty \leq \epsilon. \quad (12)$$

More formally, we need to prove that $\{\omega \in \Omega : \|\widehat{\mathbf{m}}_n(\omega) - \mathbf{m}(\omega)\|_\infty < \min(1, \delta)\} \subset \{\omega \in \Omega : \sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{G}}_{\mathbf{u}}(\omega) - \mathbf{G}_{\mathbf{u}}(\omega)\|_\infty \leq \epsilon\}$, where ω is an outcome and Ω is the sample space. For simplicity, we suppress the explicit dependence on ω in the notation. Fix $\mathbf{u} \in \mathcal{U}_n$, $s \in [0, T]$ and assume that $\|\widehat{\mathbf{m}}_n - \mathbf{m}\|_\infty < \min(1, \delta)$. The equality $\|\widehat{\mathbf{x}}_{\mathbf{u}} - \mathbf{x}_{\mathbf{u}}\|_\infty = \|\widehat{\mathbf{m}}_n - \mathbf{m}\|_\infty$ implies that $\|\widehat{\mathbf{x}}_{\mathbf{u}}(s) - \mathbf{x}_{\mathbf{u}}(s)\| < \delta$. Also, by definition of K it holds that $\mathbf{x}_{\mathbf{u}}(s) \in B_{K+1}$, and by triangle inequality we have $\widehat{\mathbf{x}}_{\mathbf{u}}(s) \in B_{K+1}$. Consequently, by uniform continuity on B_{K+1} we have

$$\|\mathbf{g}(\widehat{\mathbf{x}}_{\mathbf{u}}(s)) - \mathbf{g}(\mathbf{x}_{\mathbf{u}}(s))\| < \epsilon/T, \quad s \in [0, T].$$

Using the derived bound we obtain that for any $\mathbf{u} \in \mathcal{U}_n$

$$\|\widehat{\mathbf{G}}_{\mathbf{u}} - \mathbf{G}_{\mathbf{u}}\|_\infty \leq \sup_{[0, T]} \int_0^t \|\mathbf{g}(\widehat{\mathbf{x}}_{\mathbf{u}}(s)) - \mathbf{g}(\mathbf{x}_{\mathbf{u}}(s))\| ds \leq \int_0^T \|\mathbf{g}(\widehat{\mathbf{x}}_{\mathbf{u}}(s)) - \mathbf{g}(\mathbf{x}_{\mathbf{u}}(s))\| ds < T\epsilon/T = \epsilon.$$

Since this holds for any $\mathbf{u} \in \mathcal{U}_n$ it follows that $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\hat{\mathbf{G}}_{\mathbf{u}} - \mathbf{G}_{\mathbf{u}}\|_{\infty} \leq \epsilon$. Hence, (12) is proved. Finally, (12) and consistency of $\hat{\mathbf{m}}_n(\cdot)$ imply

$$P\left(\sup_{\mathbf{u} \in \mathcal{U}_n} \|\hat{\mathbf{G}}_{\mathbf{u}} - \mathbf{G}_{\mathbf{u}}\|_{\infty} > \epsilon\right) \leq P(\|\hat{\mathbf{m}}_n - \mathbf{m}\|_{\infty} \geq \min(1, \delta)) \rightarrow 0,$$

as $n \rightarrow \infty$, which proves the first statement. Continuity of \mathbf{g} implies its boundedness on B_{K+1} i.e. there exists $C > 0$ such that $\|\mathbf{g}(\mathbf{x})\| \leq C$ for every $\mathbf{x} \in B_{K+1}$. Fix any $\mathbf{u} \in \mathcal{U}_n$ and $s \in [0, T]$. Since $\mathbf{x}_{\mathbf{u}}(s) \in B_{K+1}$ we obtain

$$\|\mathbf{G}_{\mathbf{u}}\|_{\infty} = \sup_{t \in [0, T]} \left\| \int_0^t \mathbf{g}(\mathbf{x}_{\mathbf{u}}(s)) ds \right\| \leq \int_0^T \|\mathbf{g}(\mathbf{x}_{\mathbf{u}}(s))\| ds \leq TC,$$

which is the second claim.

(iii) and (iv) We first prove that

$$\sup_{\mathbf{u} \in \mathcal{U}_n} \|\hat{\mathbf{A}}_{\mathbf{u}} - \mathbf{A}_{\mathbf{u}}\| = o_P(1), \sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{A}_{\mathbf{u}}\| = O(1), \sup_{\mathbf{u} \in \mathcal{U}_n} \|\hat{\mathbf{A}}_{\mathbf{u}}\| = O_P(1), \quad (13)$$

$$\sup_{\mathbf{u} \in \mathcal{U}_n} \|\hat{\mathbf{B}}_{\mathbf{u}} - \mathbf{B}_{\mathbf{u}}\| = o_P(1), \sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{B}_{\mathbf{u}}\| = O(1), \sup_{\mathbf{u} \in \mathcal{U}_n} \|\hat{\mathbf{B}}_{\mathbf{u}}\| = O_P(1). \quad (14)$$

Indeed, the first assertion in (13) follows from Lemma 4.2 (iii) and the result (ii) of this lemma. The second assertion is a consequence of $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{G}_{\mathbf{u}}\|_{\infty} = O(1)$ and the inequality

$$\|\mathbf{A}_{\mathbf{u}}\| = \left\| \int_0^T \mathbf{G}_{\mathbf{u}}(t) dt \right\| \leq T \|\mathbf{G}_{\mathbf{u}}\|_{\infty}.$$

By taking into account the results (ii) of this lemma and applying Lemma 4.2 (i)-(iii) we obtain the first equality in (14). The claim $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{B}_{\mathbf{u}}\| = O(1)$ follows from $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{G}_{\mathbf{u}}\|_{\infty} = O(1)$ and the inequality

$$\|\mathbf{B}_{\mathbf{u}}\| \leq \int_0^T \|\mathbf{G}_{\mathbf{u}}^{\top}(t)\| \|\mathbf{G}_{\mathbf{u}}(t)\| dt = \int_0^T \|\mathbf{G}_{\mathbf{u}}(t)\|^2 dt \leq T \|\mathbf{G}_{\mathbf{u}}\|_{\infty}^2.$$

Finally, repeated application of Lemma 4.2 and already proven results yield (iii) and (iv). \square

Proof of Theorem 2.3. By the assumption of the theorem $\hat{\mathbf{m}}_n(\cdot)$ is a consistent estimator of $\mathbf{m}(\cdot)$. If we show that $\hat{\mathbf{u}}_n(\cdot)$ is a consistent estimator of $\mathbf{u}^*(\cdot)$ then $(\hat{\mathbf{m}}_n(\cdot), \hat{\mathbf{u}}_n(\cdot))$ is a consistent estimator of $\mathbf{x}(\cdot)$. Then by Theorem 1 of [4] $(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\xi}}_n)$ is a consistent estimator of $(\boldsymbol{\theta}, \boldsymbol{\xi})$. Consistency of $\hat{\mathbf{u}}_n(\cdot)$ is proven by verifying the conditions C3-C7 from Section 1. We have divided the proof into a sequence of lemmas. That the conditions C3-C5 are satisfied is shown in Lemma 3.2. C6 and C7 are proven in Lemmas 3.3 and 3.4, respectively. \square

Lemma 3.2. (i) *Functional $M(\mathbf{u})$ is continuous at \mathbf{u}^* in \mathcal{U} and $M(\mathbf{u}^*) < +\infty$.*

(ii) *$M_n(\mathbf{u})$ is measurable function of the data $\{Y_j(t_i)\}_{j,i}$ for all $\mathbf{u} \in \mathcal{U}_n$.*

(iii) *For any data $\{Y_j(t_i)\}_{j,i}$, $M_n(\mathbf{u})$ is lower semicontinuous on $\mathbf{u} \in \mathcal{U}_n$ under $\|\cdot\|_{\infty}$.*

Proof. (i)

We show that the mappings $\mathbf{u} \mapsto \mathbf{x}_{\mathbf{u}}$, $\mathbf{u} \mapsto \mathbf{G}_{\mathbf{u}}$, $\mathbf{u} \mapsto \mathbf{A}_{\mathbf{u}}$, $\mathbf{u} \mapsto \mathbf{B}_{\mathbf{u}}$ are continuous at \mathbf{u}^* . The result will then follow by repeated application of Lemma 4.1. Fix $\epsilon > 0$ and take $\delta = \epsilon$. Then $\|\mathbf{u} - \mathbf{u}^*\|_{\infty} < \delta$ implies

$$\|\mathbf{x}_{\mathbf{u}} - \mathbf{x}_{\mathbf{u}^*}\|_{\infty} = \|\mathbf{u} - \mathbf{u}^*\|_{\infty} < \delta = \epsilon,$$

which establishes the continuity of $\mathbf{u} \mapsto \mathbf{x}_{\mathbf{u}}$.

To prove continuity of $\mathbf{u} \mapsto \mathbf{G}_{\mathbf{u}}$ it is sufficient to prove continuity of $\mathbf{x}_{\mathbf{u}} \mapsto \mathbf{G}_{\mathbf{u}}$ because the composition of continuous maps is continuous. Fix $\epsilon > 0$. Let $K = \sup_{\mathbf{u} \in \mathcal{U}} \|\mathbf{x}_{\mathbf{u}}\|_{\infty}$. Under our assumption, the solution \mathbf{m} of the

differential equation is bounded, and also any $\mathbf{u} \in \mathcal{U}$ is bounded, hence, $\|\mathbf{x}_{\mathbf{u}}\|_{\infty} < \infty$, for any $\mathbf{u} \in \mathcal{U}$. Now we have $K < \infty$ because by (i) the mapping $\mathbf{u} \mapsto \mathbf{x}_{\mathbf{u}}$ is continuous and by Assumption 2.1 d) \mathcal{U} is compact. Continuity of $\mathbf{g}(\cdot)$ on \mathbb{R}^d implies its uniform continuity on the compact ball $B_K = \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x}\| \leq K\}$ and consequently there exists $\delta > 0$ such that for all $\mathbf{x}_1, \mathbf{x}_2 \in B_K$ with $\|\mathbf{x}_1 - \mathbf{x}_2\| < \delta$ the inequality $\|\mathbf{g}(\mathbf{x}_1) - \mathbf{g}(\mathbf{x}_2)\| < \epsilon/T$ holds. Finally, for any $\mathbf{u} \in \mathcal{U}$ such that $\|\mathbf{x}_{\mathbf{u}} - \mathbf{x}_{\mathbf{u}^*}\|_{\infty} < \delta$ we have

$$\|\mathbf{G}_{\mathbf{u}} - \mathbf{G}_{\mathbf{u}^*}\|_{\infty} \leq \sup_{[0,T]} \int_0^t \|\mathbf{g}(\mathbf{x}_{\mathbf{u}}(s)) - \mathbf{g}(\mathbf{x}_{\mathbf{u}^*}(s))\| ds \leq \int_0^T \|\mathbf{g}(\mathbf{x}_{\mathbf{u}}(s)) - \mathbf{g}(\mathbf{x}_{\mathbf{u}^*}(s))\| ds < T\epsilon/T = \epsilon.$$

Continuity of $\mathbf{x}_{\mathbf{u}} \mapsto \mathbf{G}_{\mathbf{u}}$ is proven. Continuity of $\mathbf{u} \mapsto \mathbf{A}_{\mathbf{u}}$ and $\mathbf{u} \mapsto \mathbf{B}_{\mathbf{u}}$ follows from continuity of $\mathbf{u} \mapsto \mathbf{G}_{\mathbf{u}}$ and repeated application of Lemma 4.1.

(ii)

Fix $\mathbf{u} \in \mathcal{U}_n$. The mapping $\{Y_j(t_i)\}_{j,i} \mapsto M_n(\mathbf{u})$ is measurable as a composition of measurable mappings. Indeed, by definition the estimator $\hat{\mathbf{m}}_n$ is a measurable function of the data $\{Y_j(t_i)\}_{j,i}$. Also, M_n is a measurable function of $\hat{\mathbf{m}}_n$ because it is continuous on \mathcal{M} under $\|\cdot\|_{\infty}$. The proof of the last claim is the same like the proof of continuity of M , presented in (i).

(iii)

Fix any data $\{Y_j(t_i)\}_{j,i}$. Lower semicontinuity of M_n is implied by its continuity. The mapping $\mathbf{u} \mapsto M_n(\mathbf{u})$ is indeed continuous because it has the same form as $\mathbf{u} \mapsto M(\mathbf{u})$, with the difference that \mathbf{m} is substituted with $\hat{\mathbf{m}}$. But $\hat{\mathbf{m}}$ is fixed because the data is and so the proof of continuity is the same like in (i). \square

Lemma 3.3. *For all $\epsilon > 0$, $M(\mathbf{u}^*) < \inf_{\{\mathbf{u} \in \mathcal{U} : \|\mathbf{u} - \mathbf{u}^*\| \geq \epsilon\}} M(\mathbf{u})$.*

Proof. We will prove the statement by showing that \mathbf{u}^* is a unique minimum of M . Since for any $\mathbf{u} \in \mathcal{U}$ $M(\mathbf{u}) \geq 0$ and $M(\mathbf{u}^*) = 0$ it follows that \mathbf{u}^* is a minimum of M . We now show that if $M(\mathbf{u}) = 0$ then $\mathbf{u} = \mathbf{u}^*$, which will imply that \mathbf{u}^* is the unique minimum of M . Fix $\mathbf{u} \in \mathcal{U}$ and assume $M(\mathbf{u}) = 0$. The integrand in M is nonnegative and thus equal to zero Lebesgue almost everywhere. Its continuity further implies that it must be equal to zero everywhere. This yields

$$\mathbf{x}_{\mathbf{u}}(t) - \boldsymbol{\xi}_{\mathbf{u}} - \int_0^t \mathbf{F}(\mathbf{x}_{\mathbf{u}}(s), \boldsymbol{\theta}_{\mathbf{u}}) ds = \mathbf{0}, \quad t \in [0, T],$$

where $\mathbf{0}$ is d -dimensional zero vector. From the previous display we obtain that $\mathbf{x}'_{\mathbf{u}}(t) = \mathbf{F}(\mathbf{x}_{\mathbf{u}}(t), \boldsymbol{\theta}_{\mathbf{u}})$, $t \in [0, T]$ and $\mathbf{x}_{\mathbf{u}}(0) = \boldsymbol{\xi}_{\mathbf{u}}$, which implies that $\mathbf{x}_{\mathbf{u}}(\cdot)$ is a solution of the system of ODEs

$$\begin{cases} \mathbf{x}'(t) = \mathbf{F}(\mathbf{x}(t), \boldsymbol{\theta}_{\mathbf{u}}), & t \in [0, T], \\ \mathbf{x}(0) = \boldsymbol{\xi}_{\mathbf{u}}. \end{cases} \quad (15)$$

But according to Assumption 2.1 a) the solution $\mathbf{x}(\cdot, \boldsymbol{\theta}_{\mathbf{u}}, \boldsymbol{\xi}_{\mathbf{u}})$ of the ODE system (15) is unique, so we must have $\mathbf{x}_{\mathbf{u}}(\cdot) = \mathbf{x}(\cdot, \boldsymbol{\theta}_{\mathbf{u}}, \boldsymbol{\xi}_{\mathbf{u}})$. This is equivalent to $\mathbf{m}(\cdot, \boldsymbol{\theta}, \boldsymbol{\xi}) = \mathbf{m}(\cdot, \boldsymbol{\theta}_{\mathbf{u}}, \boldsymbol{\xi}_{\mathbf{u}})$ and $\mathbf{u}(\cdot) = \mathbf{u}(\cdot, \boldsymbol{\theta}_{\mathbf{u}}, \boldsymbol{\xi}_{\mathbf{u}})$. By Remark 2.2 it follows that $(\boldsymbol{\theta}, \boldsymbol{\xi}) = (\boldsymbol{\theta}_{\mathbf{u}}, \boldsymbol{\xi}_{\mathbf{u}})$, which in turn, by Assumption 2.1 a), implies $\mathbf{x}(\cdot, \boldsymbol{\theta}, \boldsymbol{\xi}) = \mathbf{x}(\cdot, \boldsymbol{\theta}_{\mathbf{u}}, \boldsymbol{\xi}_{\mathbf{u}})$. Finally, from the last equality we have $\mathbf{u}(\cdot, \boldsymbol{\theta}, \boldsymbol{\xi}) = \mathbf{u}(\cdot, \boldsymbol{\theta}_{\mathbf{u}}, \boldsymbol{\xi}_{\mathbf{u}})$, i.e. $\mathbf{u}^*(\cdot) = \mathbf{u}(\cdot)$. This is the desired conclusion. \square

Lemma 3.4. $\sup_{\mathbf{u} \in \mathcal{U}_n} |M_n(\mathbf{u}) - M(\mathbf{u})| \xrightarrow{P} 0$, as $n \rightarrow \infty$.

Proof. We follow the idea of the proof of Proposition 3.2 of [6]. Inequality $||a||^2 - ||b||^2| \leq ||a - b||(||a|| + ||b||)$, and Cauchy Shwartz and triangle inequalities in $L^2[0, T]$ imply

$$\begin{aligned} |M_n(\mathbf{u}) - M(\mathbf{u})| &= \left| \int_0^T \left(\|\widehat{\mathbf{x}}_{\mathbf{u}}(t) - \widehat{\boldsymbol{\xi}}_{\mathbf{u}} - \widehat{\mathbf{G}}_{\mathbf{u}}(t)\widehat{\boldsymbol{\theta}}_{\mathbf{u}}\|^2 - \|\mathbf{x}_{\mathbf{u}}(t) - \boldsymbol{\xi}_{\mathbf{u}} - \mathbf{G}_{\mathbf{u}}(t)\boldsymbol{\theta}_{\mathbf{u}}\|^2 \right) dt \right| \\ &\leq \sqrt{\int_0^T \|\widehat{\mathbf{x}}_{\mathbf{u}}(t) - \mathbf{x}_{\mathbf{u}}(t) - \widehat{\boldsymbol{\xi}}_{\mathbf{u}} + \boldsymbol{\xi}_{\mathbf{u}} - \widehat{\mathbf{G}}_{\mathbf{u}}(t)\widehat{\boldsymbol{\theta}}_{\mathbf{u}} + \mathbf{G}_{\mathbf{u}}(t)\boldsymbol{\theta}_{\mathbf{u}}\|^2 dt} \\ &\times \left\{ \sqrt{\int_0^T \|\widehat{\mathbf{x}}_{\mathbf{u}}(t) - \widehat{\boldsymbol{\xi}}_{\mathbf{u}} - \widehat{\mathbf{G}}_{\mathbf{u}}(t)\widehat{\boldsymbol{\theta}}_{\mathbf{u}}\|^2 dt} + \sqrt{\int_0^T \|\mathbf{x}_{\mathbf{u}}(t) - \boldsymbol{\xi}_{\mathbf{u}} - \mathbf{G}_{\mathbf{u}}(t)\boldsymbol{\theta}_{\mathbf{u}}\|^2 dt} \right\} \\ &:= \sqrt{T_1(\mathbf{u})} \{ \sqrt{T_2(\mathbf{u})} + \sqrt{T_3(\mathbf{u})} \}. \end{aligned}$$

Since \mathcal{U}_n is compact we have $\sup_{\mathbf{u} \in \mathcal{U}_n} T_3(\mathbf{u}) = O(1)$. Results of Lemma 3.1 and repeated use of Lemma 4.2 show that $\sup_{\mathbf{u} \in \mathcal{U}_n} T_1(\mathbf{u}) = o_P(1)$. Finally, from triangle inequality and inequality $(a + b)^2 \leq 2(a^2 + b^2)$ it follows that

$$\begin{aligned} T_2(\mathbf{u}) &= \int_0^T \|\widehat{\mathbf{x}}_{\mathbf{u}}(t) - \widehat{\boldsymbol{\xi}}_{\mathbf{u}} - \widehat{\mathbf{G}}_{\mathbf{u}}(t)\widehat{\boldsymbol{\theta}}_{\mathbf{u}}\|^2 dt \\ &\leq 2 \int_0^T \|\widehat{\mathbf{x}}_{\mathbf{u}}(t) - \mathbf{x}_{\mathbf{u}}(t) - \widehat{\boldsymbol{\xi}}_{\mathbf{u}} + \boldsymbol{\xi}_{\mathbf{u}} - \widehat{\mathbf{G}}_{\mathbf{u}}(t)\widehat{\boldsymbol{\theta}}_{\mathbf{u}} + \mathbf{G}_{\mathbf{u}}(t)\boldsymbol{\theta}_{\mathbf{u}}\|^2 dt + 2 \int_0^T \|\mathbf{x}_{\mathbf{u}}(t) - \boldsymbol{\xi}_{\mathbf{u}} - \mathbf{G}_{\mathbf{u}}(t)\boldsymbol{\theta}_{\mathbf{u}}\|^2 dt \\ &= 2T_1(\mathbf{u}) + 2T_3(\mathbf{u}), \end{aligned}$$

whence $\sup_{\mathbf{u} \in \mathcal{U}_n} T_2(\mathbf{u}) = O_P(1)$. This completes the proof. \square

4. APPENDIX A

In what follows, \mathcal{U}_n and \mathcal{U} satisfy conditions from previous sections. Recall that for a matrix function $\mathbf{M} : [0, T] \rightarrow \mathbb{R}^{m \times p}$ we use the norm $\|\mathbf{M}\|_{\infty} = \sup_{t \in [0, T]} \|\mathbf{M}(t)\|$, where $\|\cdot\|$ is the Frobenius norm on $\mathbb{R}^{m \times p}$. Also, $\widehat{\mathbf{M}}$ denotes the estimator of \mathbf{M} based on the data $\{Y_j(t_i)\}_{j,i}$.

Lemma 4.1. *For $\mathbf{u} \in \mathcal{U}_n$ (\mathcal{U}), let $\mathbf{M}_{\mathbf{u}} : [0, T] \rightarrow \mathbb{R}^{m \times p}$, $\mathbf{P}_{\mathbf{u}} : [0, T] \rightarrow \mathbb{R}^{p \times s}$ and $\mathbf{Q}_{\mathbf{u}} \in \mathbb{R}^{m \times m}$. If the mappings $\mathbf{u} \mapsto \mathbf{M}_{\mathbf{u}}$, $\mathbf{u} \mapsto \mathbf{P}_{\mathbf{u}}$, $\mathbf{u} \mapsto \mathbf{Q}_{\mathbf{u}}$ are continuous on \mathcal{U}_n (\mathcal{U}) then so are $\mathbf{u} \mapsto \mathbf{M}_{\mathbf{u}}^{\top}$, $\mathbf{u} \mapsto \mathbf{M}_{\mathbf{u}}\mathbf{P}_{\mathbf{u}}$, $\mathbf{u} \mapsto \int_0^T \mathbf{M}_{\mathbf{u}}(t) dt$, $\mathbf{u} \mapsto \mathbf{Q}_{\mathbf{u}}^{-1}$.*

Proof. Fix $\mathbf{u}_0 \in \mathcal{U}_n(\mathcal{U})$. Continuity of the mappings $\mathbf{u} \mapsto \mathbf{M}_{\mathbf{u}}^{\top}$, $\mathbf{u} \mapsto \mathbf{M}_{\mathbf{u}}\mathbf{P}_{\mathbf{u}}$, $\mathbf{u} \mapsto \int_0^T \mathbf{M}_{\mathbf{u}}(t) dt$ at \mathbf{u}_0 follows from

$$\begin{aligned} \|\mathbf{M}_{\mathbf{u}}^{\top} - \mathbf{M}_{\mathbf{u}_0}^{\top}\|_{\infty} &= \|\mathbf{M}_{\mathbf{u}} - \mathbf{M}_{\mathbf{u}_0}\|_{\infty}, \\ \|\mathbf{M}_{\mathbf{u}}\mathbf{P}_{\mathbf{u}} - \mathbf{M}_{\mathbf{u}_0}\mathbf{P}_{\mathbf{u}_0}\|_{\infty} &\leq \|\mathbf{M}_{\mathbf{u}}\|_{\infty} \|\mathbf{P}_{\mathbf{u}} - \mathbf{P}_{\mathbf{u}_0}\|_{\infty} + \|\mathbf{M}_{\mathbf{u}} - \mathbf{M}_{\mathbf{u}_0}\|_{\infty} \|\mathbf{P}_{\mathbf{u}_0}\|_{\infty}, \\ \left\| \int_0^T \mathbf{M}_{\mathbf{u}}(t) dt - \int_0^T \mathbf{M}_{\mathbf{u}_0}(t) dt \right\| &\leq T \|\mathbf{M}_{\mathbf{u}} - \mathbf{M}_{\mathbf{u}_0}\|_{\infty}, \end{aligned} \tag{16}$$

and continuity of the mappings $\mathbf{u} \mapsto \mathbf{M}_{\mathbf{u}}$ and $\mathbf{u} \mapsto \mathbf{P}_{\mathbf{u}}$. Continuity of $\mathbf{u} \mapsto \mathbf{Q}_{\mathbf{u}}^{-1}$ follows from continuity of $\mathbf{u} \mapsto \mathbf{Q}_{\mathbf{u}}$ and continuity of the matrix inversion. \square

Lemma 4.2. *For $\mathbf{u} \in \mathcal{U}_n$ let $\mathbf{M}_{\mathbf{u}}, \widehat{\mathbf{M}}_{\mathbf{u}} : [0, T] \rightarrow \mathbb{R}^{m \times p}$, $\mathbf{P}_{\mathbf{u}}, \widehat{\mathbf{P}}_{\mathbf{u}} : [0, T] \rightarrow \mathbb{R}^{p \times s}$ and $\mathbf{Q}_{\mathbf{u}} \in \mathbb{R}^{m \times m}$ be such that*

$$\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{M}}_{\mathbf{u}} - \mathbf{M}_{\mathbf{u}}\|_{\infty} = o_P(1), \quad \sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{P}}_{\mathbf{u}} - \mathbf{P}_{\mathbf{u}}\|_{\infty} = o_P(1), \quad \sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{Q}}_{\mathbf{u}} - \mathbf{Q}_{\mathbf{u}}\| = o_P(1).$$

(i) If $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{M}_{\mathbf{u}}\|_{\infty} = O(1)$ then as $n \rightarrow \infty$

$$\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{M}}_{\mathbf{u}}^{\top} - \mathbf{M}_{\mathbf{u}}^{\top}\|_{\infty} = o_P(1), \quad \sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{M}_{\mathbf{u}}^{\top}\|_{\infty} = O(1), \quad \sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{M}}_{\mathbf{u}}^{\top}\|_{\infty} = O_P(1).$$

(ii) If $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{M}_{\mathbf{u}}\|_{\infty} = O(1)$ and $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{P}}_{\mathbf{u}}\|_{\infty} = O_P(1)$ then as $n \rightarrow \infty$

$$\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{M}}_{\mathbf{u}} \widehat{\mathbf{P}}_{\mathbf{u}} - \mathbf{M}_{\mathbf{u}} \mathbf{P}_{\mathbf{u}}\|_{\infty} = o_P(1), \quad \sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{M}}_{\mathbf{u}} \widehat{\mathbf{P}}_{\mathbf{u}}\|_{\infty} = O_P(1).$$

(iii) $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\int_0^T \widehat{\mathbf{M}}_{\mathbf{u}}(t) dt - \int_0^T \mathbf{M}_{\mathbf{u}}(t) dt\| = o_P(1)$, as $n \rightarrow \infty$.

(iv) If $\sup_{\mathbf{u} \in \mathcal{U}} \|\mathbf{Q}_{\mathbf{u}}\| < +\infty$ then as $n \rightarrow \infty$

$$\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{Q}}_{\mathbf{u}}^{-1} - \mathbf{Q}_{\mathbf{u}}^{-1}\| = o_P(1), \quad \sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{Q}_{\mathbf{u}}^{-1}\| = O(1), \quad \sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{Q}}_{\mathbf{u}}^{-1}\| = O_P(1).$$

Proof. The assertions regarding the boundedness in probability in (i)-(iv) follow from other results by using triangle inequality. As for the statements regarding boundedness we have $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{M}_{\mathbf{u}}^{\top}\|_{\infty} = \sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{M}_{\mathbf{u}}\|_{\infty} = O(1)$. Also, by Lemma 4.1 the mapping $\mathbf{u} \mapsto \mathbf{Q}_{\mathbf{u}}^{-1}$ is continuous on \mathcal{U}_n and thus bounded. This proves $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{Q}_{\mathbf{u}}^{-1}\|_{\infty} = O(1)$.

Now we prove the statements regarding little-O in probability. Replacing $\mathbf{M}_{\mathbf{u}_0}$ and $\mathbf{P}_{\mathbf{u}_0}$ by $\widehat{\mathbf{M}}_{\mathbf{u}}$ and $\widehat{\mathbf{P}}_{\mathbf{u}}$ in (16) yields inequalities which imply (i),(ii) and (iii).

To prove (iv) introduce $K = \sup_{\mathbf{u} \in \mathcal{U}} \|\mathbf{Q}_{\mathbf{u}}\|$, which is by Assumption 2.1 d) finite. Continuity of the matrix inverse on the closed compact ball $B_{K+1} = \{\mathbf{Q} \in \mathbb{R}^{m \times m} : \|\mathbf{Q}\| \leq K+1\}$ implies its uniform continuity on B_{K+1} . Fix $\epsilon > 0$. There exists $\delta > 0$ such that for any $\mathbf{Q}_1, \mathbf{Q}_2 \in B_{K+1}$ inequality $\|\mathbf{Q}_1 - \mathbf{Q}_2\| < \delta$ implies $\|\mathbf{Q}_1^{-1} - \mathbf{Q}_2^{-1}\| < \epsilon$. We will show that

$$\sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{Q}_{\mathbf{u}} - \widehat{\mathbf{Q}}_{\mathbf{u}}\|_{\infty} < \min(1, \delta) \Rightarrow \sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{Q}}_{\mathbf{u}}^{-1} - \mathbf{Q}_{\mathbf{u}}^{-1}\| \leq \epsilon. \quad (17)$$

As in proof of Lemma 3.1 (ii), here we also suppress the explicit dependence on the outcome ω in the notation. Inequality $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{Q}_{\mathbf{u}} - \widehat{\mathbf{Q}}_{\mathbf{u}}\|_{\infty} < \min(1, \delta)$ implies that for any $\mathbf{u} \in \mathcal{U}_n$ it holds $\|\mathbf{Q}_{\mathbf{u}} - \widehat{\mathbf{Q}}_{\mathbf{u}}\| < \delta$. By definition of K it holds that $\mathbf{Q}_{\mathbf{u}} \in B_{K+1}$, and by triangle inequality we obtain $\widehat{\mathbf{Q}}_{\mathbf{u}} \in B_{K+1}$. Uniform continuity now implies that $\|\widehat{\mathbf{Q}}_{\mathbf{u}}^{-1} - \mathbf{Q}_{\mathbf{u}}^{-1}\| < \epsilon$. Since this holds for any $\mathbf{u} \in \mathcal{U}_n$ it follows that $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{Q}}_{\mathbf{u}}^{-1} - \mathbf{Q}_{\mathbf{u}}^{-1}\| \leq \epsilon$. Therefore, (17) is proved. Finally, (17) and the assumption $\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{Q}}_{\mathbf{u}} - \mathbf{Q}_{\mathbf{u}}\| = o_P(1)$ imply

$$P(\sup_{\mathbf{u} \in \mathcal{U}_n} \|\widehat{\mathbf{Q}}_{\mathbf{u}}^{-1} - \mathbf{Q}_{\mathbf{u}}^{-1}\| > \epsilon) \leq P(\sup_{\mathbf{u} \in \mathcal{U}_n} \|\mathbf{Q}_{\mathbf{u}} - \widehat{\mathbf{Q}}_{\mathbf{u}}\|_{\infty} \geq \min(1, \delta)) \rightarrow 0,$$

as $n \rightarrow \infty$, which is the first assertion. □

Lemma 4.3. (i) Let $\mathbf{f} : [0, T] \rightarrow \mathbb{R}^m$ be a vector-valued function in $L^1[0, T]$. Then

$$\left\| \int_0^T \mathbf{f}(t) dt \right\| \leq \int_0^T \|\mathbf{f}(t)\| dt.$$

(ii) Let $\mathbf{M} \in \mathbb{R}^{m \times p}$ and $\mathbf{P} \in \mathbb{R}^{p \times s}$. Then

$$\|\mathbf{MP}\| \leq \|\mathbf{M}\| \|\mathbf{P}\|.$$

Proof. For (i) see page 540 of [7] and for (ii) page 550 of [1]. □

The first author is thankful to Bartek Knapik, Shota Gugushvili and Eduard Belitser for useful discussions.

REFERENCES

- [1] Dennis S Bernstein. *Matrix mathematics: theory, facts, and formulas*. Princeton University Press, 2009.
- [2] Xiaohong Chen. Large sample sieve estimation of semi-nonparametric models. *Handbook of econometrics*, 6:5549–5632, 2007.
- [3] Itai Dattner. A model-based initial guess for estimating parameters in systems of ordinary differential equations. *Biometrics*, 71(4):1176–1184, 2015.
- [4] Itai Dattner and Chris A J Klaassen. Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters. *Electronic Journal of Statistics*, 9(2):1939–1973, 2015.
- [5] Ying Ding and Bin Nan. A sieve m-theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data. *Annals of statistics*, 39(6):2795, 2011.
- [6] Shota Gugushvili, Chris AJ Klaassen, et al. \sqrt{n} -consistent parameter estimation for systems of ordinary differential equations: bypassing numerical integration via smoothing. *Bernoulli*, 18(3):1061–1098, 2012.
- [7] Frank Jones. *Lebesgue integration on Euclidean space*. Jones & Bartlett Learning, 2001.
- [8] Xin Qi and Hongyu Zhao. Asymptotic efficiency and finite-sample properties of the generalized profiling estimation of parameters in ordinary differential equations. *The Annals of Statistics*, 38(1):435–481, 2010.
- [9] Jim O Ramsay, G Hooker, D Campbell, and J Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5):741–796, 2007.
- [10] Ivan Vujačić, Itai Dattner, Javier González, and Ernst Wit. Time-course window estimator for ordinary differential equations linear in the parameters. *Statistics and Computing*, 25(6):1057–1070, 2015.
- [11] Halbert White and J Wooldridge. Some results on sieve estimation with dependent observations. *Nonparametric and Semiparametric Methods in Economics*, pages 459–493, 1991.